

AD-A187 118

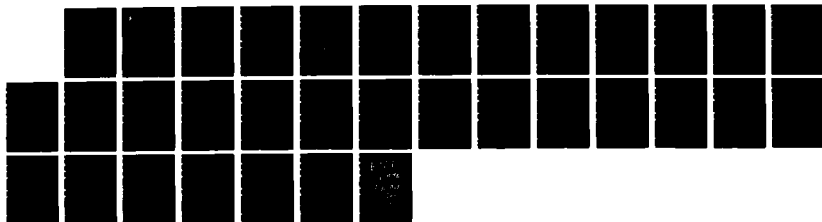
MEASURING MENTAL WORKLOAD: A PERFORMANCE BATTERY(U)
HUMAN ENGINEERING LAB ABERDEEN PROVING GROUND MD
L A WHITAKER ET AL. SEP 87 HEL-TN-21-87

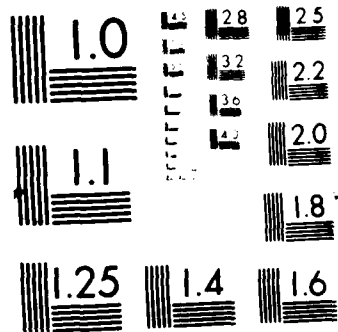
1/1

UNCLASSIFIED

F/G 5/9

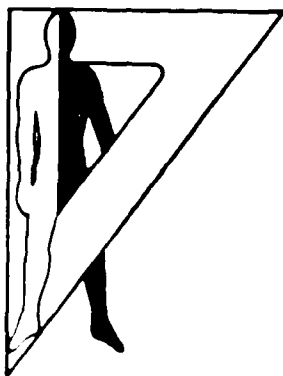
NL





AD-A187 118

4



AD

DTIC FILE COPY

Technical Memorandum 21-87

MEASURING MENTAL WORKLOAD: A PERFORMANCE BATTERY

Leslie A. Whitaker
Klein Associates

Lynn C. Oatman
Human Engineering Laboratory

Matthew D. Shank
University of Missouri-St. Louis

September 1987
AMCMS Code 611102.74A0011

Approved for public release;
distribution is unlimited.

U. S. ARMY HUMAN ENGINEERING LABORATORY

Aberdeen Proving Ground, Maryland

DTIC
ELECTE
DEC 09 1987
S E D

©Apple IIe, Applesoft BASIC, Apple Imagewriter, and DOS are registered trademarks of Apple Computer, Inc.

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial products.

ADA187118

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp Date Jun 30, 1986

1a REPORT SECURITY CLASSIFICATION <u>Unclassified</u>		1b RESTRICTIVE MARKINGS	
2a SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b DECLASSIFICATION/DOWNGRADING SCHEDULE		5 MONITORING ORGANIZATION REPORT NUMBER(S)	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Memorandum 21-87		7a NAME OF MONITORING ORGANIZATION	
6a NAME OF PERFORMING ORGANIZATION Human Engineering Laboratory	6b OFFICE SYMBOL (If applicable) SLCME	7b ADDRESS (City, State, and ZIP Code)	
6c ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21005-5001		9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8a NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	10. SOURCE OF FUNDING NUMBERS	
8c ADDRESS (City, State, and ZIP Code)		PROGRAM ELEMENT NO. 6.11.02.A	PROJECT NO. 1L161102B74A
		TASK NO.	WORK UNIT ACCESSION NO.
11 TITLE (Include Security Classification) MEASURING MENTAL WORKLOAD: A PERFORMANCE BATTERY			
12 PERSONAL AUTHOR(S) Leslie A. Whitaker Matthew D. Shank Lynn C. Oatman			
13a TYPE OF REPORT Final	13b TIME COVERED FROM _____ TO _____	14 DATE OF REPORT (Year, Month, Day) September 1987	15 PAGE COUNT 29
16 SUPPLEMENTARY NOTATION			
17 COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
05	08		
		.. mental workload visual search mental arithmetic memory search assessment battery	
19 ABSTRACT (Continue on reverse if necessary and identify by block number) A performance test battery consisting of three tasks (memory/visual search, serial arithmetic, and column addition) was developed for use on an Apple® IIe microcomputer. For each task, several workload parameters were varied via a menu. In three separate experiments, performance variables (response latency and error rate) were found to be sensitive to changes in workload. The test battery is an easy method of varying workload through the use of a controlled laboratory task. Future research measuring subjective opinions and evoked potentials is planned using this battery of tasks.			
20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21 ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL Patricia Streett		22b TELEPHONE (Include Area Code) 301-278-4478	22c OFFICE SYMBOL SLCME-SS-TS

MEASURING MENTAL WORKLOAD: A PERFORMANCE BATTERY



Leslie A. Whitaker
Klein Associates
P.O. Box 264
Yellow Springs, Ohio 45387

Lynn C. Oatman
Human Engineering Laboratory

Matthew D. Shank
University of Missouri-St. Louis

September 1987

Mission For	
THIS GRAB	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

APPROVED: _____

JOHN D. WEISZ
Director

U.S. Army Human Engineering Laboratory

Approved for public release;
distribution is unlimited.

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005-5001

PREFACE

This manuscript was written while the first author was a Battelle Summer Associate at the Human Engineering Laboratory, Aberdeen Proving Ground, Maryland.

The authors wish to thank Dr. David Thorne for his extensive help during the modification of the PAB software.

CONTENTS

EXPERIMENT I	7
Methods	7
Results	7
EXPERIMENT II	9
Methods	11
Results	12
EXPERIMENT III	15
Methods	15
Results	17
GENERAL DISCUSSION	21
REFERENCES	23
APPENDIX	
Analysis of Variance Tables	25
FIGURES	
1. Memory/Visual Search Task: Mean Response Latency as a Function of Memory Set Size and Target Search Set Size	8
2. Memory/Visual Search Task: Error Rates as a Function of Memory Set Size and Target Search Set Size	9
3. Add/Subtract Task: Mean Response Latency as a Function of Number of Stimulus Items and the Stimulus Durations	13
4. Add/Subtract Task: Error Rates as a Function of Number of Stimulus Items and the Stimulus Durations	14
5. Column Addition Task: Mean Response Latency as a Function of Number of Digits per Item and Number of Items to be Added	18
6. Column Addition Task: Error Rates as a Function of Number of Digits per Item and Number of Items to be Added	19
TABLES	
1. Regression Coefficients for the Memory Search Task: Predicting Response Latency (CRT) From Task Variables	10
2. Regression Coefficients for the Serial Arithmetic Task: Predicting Response Latency (CRT) From Task Variables	16
3. Regression Coefficients for the Addition Task: Predicting Response Latency (CRT) From Task Variables	20

MEASURING MENTAL WORKLOAD: A PERFORMANCE BATTERY

The measurement of mental workload has been investigated extensively during the past decade (see Wickens & Kramer, 1987, for review). There is no single definition of workload that has been found to be acceptable and, therefore, it is not surprising that there is still considerable controversy about the appropriate methods to be used in assessing workload. Despite Knowles (1963) recommendation that efforts be made to arrive at a few generally accepted tests, this goal has not been achieved.

The goals of any such method are to show sensitivity to changes in workload, lack of intrusion on the performance of the primary task, and representativeness across job categories. Chiles and Alluisi (1979) have recommended the use of laboratory tasks whenever feasible to achieve these goals. They argued that laboratory tasks are preferable to measuring the job itself (primary task measurement) because findings from one primary task do not generalize to other dissimilar jobs. This lack of representativeness is a major stumbling block in the development of workload measurements based on primary tasks. The use of a single representative task to measure generic mental workload is probably an impossible goal because of the large domain of cognitive activities found in various types of mental work (LePlat, 1978). In contrast to primary task measurements, laboratory tasks can sample specific cognitive activities and can be combined with various primary tasks to index their workload. Knowles (1963) recommended that these tasks should sample various cognitive functions found in the workplace (e.g., mental arithmetic is representative of the general problem-solving requirements of mental work). Gopher and Sanders (1984) have echoed this same recommendation 20 years later. However, there still appears to be no resolution in the final choice of measurement.

The use of more than one type of measurement has been recommended by various investigators because of the inconsistency or insensitivity of any one measurement (LePlat, 1978; Moray, 1982; Simmons, 1979). When convergent measurements from various techniques can point to specific tasks as having high workload, the evidence is far stronger than when only a single measure is used.

Moray (1982) noted the generally poor correlations among subjective opinion (operator reports of task difficulty), physiological, and performance measurements. He recommended further research attempts that assessed workload by using several measurement techniques drawn from the subjective, physiological, and performance domains. This is consistent with Wickens and Kramer's (1985) position that workload is best viewed "as a multidimensional construct that includes behavioral, physiological, and subjective aspects." (p. 316).

The present research was the first step in the development of a test battery that could compare workload measures in these three categories. The initial work was the construction of a computer program that could present generally accepted cognitive tasks at a variety of difficulty (workload) levels. These tasks were then tested to determine the influence of such difficulty manipulations on performance. The goal of these experiments was to determine a range within which performance was sensitive to changes in task parameters. On the basis of these results, subsequent experiments will compare a subset of these task conditions to measure subjective and physiological responses, as well as performance. This test of convergent measures will provide the correlations recommended by Moray (1982) to

assess task variables sensitive to all three measurement techniques. This report was written to make available the test battery and these initial performance data.

The test battery itself was based on a performance assessment battery developed by Walter Reed Biomedical Laboratory (Thorne, Genser, Sing, & Hegge, 1985). In its original form the battery consists of 13 performance tasks; however, the parameters (workload) exist at only one level for each task. In the modified program, it is possible to vary a number of stimulus parameters to determine their effect on the performance output.

In a series of three experiments, a visual search and two types of mental arithmetic tasks were examined to determine which stimulus parameters affected performance. Variation in workload levels was operationally defined as those changes in task parameters that produce changes in performance.

EXPERIMENT 1

Immediate memory for letters has been examined in a variety of paradigms to determine the effects of memory load (Sternberg, 1969) and visual search complexity (Neisser, Novick, & Lazar, 1963) on processing time. Kaplan, Carvellas, and Metlay (1966) combined both these variables in a memory and search task (MAST). While subjects completed a paper and pencil test, their eye movements were measured to determine the time spent scanning a 10-item search string of letters for an embedded subset of 1 to 10 letters. The memory set was changed for each trial; therefore, subjects were working on an immediate (short-term memory) task. Scan time increased linearly as a function of memory set size.

Folkard, Knauth, Monk, and Rutenfranz (1976) examined a similar task in which subjects searched a 20-letter string for a memorized subset of two, four, or six items; however, the memory set items were constant for an entire block of 96 trials. The authors reported that the number of lines searched varied as a function of the time of day. When the average performance at each memory set size is computed, the slope is not linear. Instead the processing time per memory item decreased as a function of memory set size.

In both these studies, subjects were allowed to return to the memory set to review it after they had begun scanning the search string. Kaplan et al. (1966) measured the time subjects spent reviewing the memory set and found it to be a positively accelerated function of memory set size. Therefore, the amount of processing time as a function of memory load is not linear when the total time spent on scanning the search string and returning to review the memory set are combined. Instead Kaplan et al. (1966) found that the total processing time per memory item increased as memory set size increased. Response latency results of the present study will be analyzed to determine whether the best fit regression function shows an increasing or decreasing time cost per memory set item.

Kaplan et al. (1966) described their task as an immediate memory task. Such tasks have been shown to be sensitive to the amount of time (retention interval) following the presentation of the memory set before recall is allowed (Peterson & Peterson, 1959). Keppel and Underwood (1962) reported that tasks with variable memory sets (i.e., a new memory set for each trial) were very susceptible to forgetting as a function of retention interval length. These results were

obtained for tasks in which the retention interval was filled with a distracting task to prevent rehearsal of the memory set. Klatzky (1984) has reviewed immediate memory processes and concluded that the retention interval can serve as a period of consolidation to develop richer memory traces that resist forgetting, if the retention interval is free of a distracting task.

The present experiment was designed to replicate the conditions tested by Kaplan et al. (1966) and Folkard et al. (1976) using a computer presentation that prevented the subject's looking back at the memory set after beginning to scan the search string. In this way, response latency would reflect the total processing time once the scan had begun. The purpose was to determine whether the search for a larger set of letters was more or less efficient (time per item) than was the search for a smaller set of memorized items. Briggs (1974) surveyed the literature to determine the shape of the function that predicted response latency from memory set size. He concluded that some data were better fit by linear (equal time per item) and some by logarithmic (less time per item as set size increases) functions. Therefore, analyses of the present data were planned to test the fit of both functions. In addition, the retention interval was varied to determine whether time to consolidate the memory set would facilitate search; hence a new memory set was used on each trial.

Each of these variables represents a different type of load in operational tasks. The memory component is critical to any coding or transcription task and has been found to be a function of both the memory set size and retention interval (e.g., Sternberg, 1969; Wickens, 1984).

METHODS

Subjects. Twenty university students were tested. Twelve of these participants were females and eight were males. Eight were graduate students, while the remaining students were undergraduates. They were found to have normal or corrected to normal visual acuity. Subjects were randomly divided into two retention interval groups with the restriction that half of the graduate students and half of the males were in each group. One group participated in the condition with a 3-second retention interval between the offset of the memory string and the onset of the search string. The other group had a 0-second retention interval. Each subject participated in a single session during which he or she completed all repeated-measures conditions in the experiment.

Design. The task levels were the following: (2, 4, or 6 memory letters) x (9 or 19 search letters). Each of these six conditions was presented in two separate blocks of 11 trials. The order of presentation of the six conditions was counterbalanced to control zero- and first-order position effects. The second block of six conditions was in the reversed order; thus providing an ABBA design.

Procedure. Each subject was tested individually. The subject was first tested to determine that his/her vision was either normal (20/25) or corrected to that level by the subject's own glasses or contact lens. The program is designed to provide a menu and prompts that allow the subject to serve as experimenter by keying in the parameters for the next block of trials; however, an experimenter was

always present in the room to answer questions and to determine whether the subject was following the correct procedure. The subject was seated at 40 cm from the screen. The memory string spanned 2.6° to 4° visual angle, while the target string spanned from 6° to 12.7° .

On each trial, a fixation point was marked by a "+" in the center of the screen. After a 1.5-second warning interval, the memory set of two, four, or six letters (centered horizontally about the fixation position) replaced the "+". After 2 seconds, the screen was blanked and either the search string appeared immediately (0-second retention group) or a 3-second retention interval occurred before the search string appeared. The search string contained either 2 or 4 letters. For either search string length, half of the strings contained all the memory set letters. The subjects responded to these sets by pressing the S (same) key. On the remaining sets, at least one of the memory set letters was missing. In these trials, the subject pressed the D (different) key.

Subjects were instructed to work as quickly as possible without making errors. The subject first completed a block of practice trials for each of the six conditions. There were five trials in each practice block. If there were no questions, the subject then began the data trials. After completing one block of 11 trials for each of the six conditions in the designated order, the subject had a 5-minute break and then completed one more block of each condition in the reverse order. Summary response time and accuracy data were displayed on the screen at the end of each block of trials. The subjects were encouraged to compete with their own scores to improve their performance.

Equipment. An Apple[®] IIe system with two disk drives, 128 K memory, and a monochromatic screen was used to present the stimulus information and register the subject's response. A timer board from California Computer Systems (Model # 7220A) was inserted in slot #2 to time the duration of the response intervals. The data were output to an Apple[®] Imagewriter printer for later analysis. The parameters and subject identification code were printed for each block of trials. The accuracy and response time for each trial were recorded as well as summary data for each 11-trial block.

The test battery itself was modified from the Walter Reed Performance Assessment Battery (Thorne et al., 1985). This program is written in Applesoft[®] BASIC under DOS[®] 3.3 and is available in both an uncompiled and compiled version. The modified test battery provided the opportunity to vary (via menu) the following parameters in the memory search task used in the present study: the number of trials in any block, the number of letters in the memory set, retention interval between the memory set and the search string, and the number of letters in the search string.

RESULTS

The mean response times to correct trials in each block were analyzed in an analysis of variance. A mixed-factors design was used to determine the significance of the grouping variable (retention interval) and the three repeated measures (memory set size, search set size, and trial block). The summary source table for the analysis of variance is included in the Appendix. The effect of

blocks was not found to be significant for main effect or interaction; therefore, the data were graphed for the entire set of blocks of trials in each condition.

The mean response times for correct responses for each group are in Table 1. The effects of memory set size, $F(2,36) = 14.78, p < .001$, and search set size, $F(2,36) = 14.20, p < .001$, Memory \times Search, $F(2,36) = 14.78, p < .001$, were all significant. The effect of retention interval was not significant for response times, but it did interact with the size of the memory set, $F(2,36) = 14.20, p < .001$. The increase in response latency due to memory set size was larger when the retention interval was 3 seconds than it was when the subjects could retrieve the target string immediately (0-second retention) (see Figure 1).

Error rates increased with increasing memory set size but were not dramatically different for the two search string sizes (9 vs. 19). It is premature to interpret the three-way interaction of memory set size, search set size, and retention interval; however, the means of this interaction are graphed in Figure 2.

When the graphs of the two dependent variables are considered together, it appears that a speed-accuracy trade-off has occurred between retention intervals. At the shorter (0-second) interval, response times are less affected by memory set size, while error rates increase sharply. At the longer (3-second) interval, the response time slope as a function of memory set size is steeper; however, error rates increased less than for the shorter retention interval. There is no apparent reason in the experimental procedure for this trade-off; therefore, this phenomenon may merit further study.

In an exploratory regression analysis, the data for latencies of correct responses were fit to linear, quadratic, and logarithmic models for each retention interval group. A "best" model was found by allowing variables to enter the equation in a forward stepwise regression (SPSS, 1970) until the addition of more variables did not significantly increase the proportion of variance explained (R^2). The best linear model entered both memory set size and search set size for each group. Both the quadratic and the logarithmic model provide a similar fit, but may be less straightforward to interpret. The regression equations for each model are listed in Table 1. None of these models has been cross-validated; however, the percent variance explained is in the range Nunnally (1975) describes as a moderately ($R = .40-.50$) or a very ($R = .50-.60$) satisfactory predictor.

None of these models fits the $RI = 0$ data as well as it fits the $RI = 3$ data. The linear model fits as well as the more complex models for each dependent variable. However, it should be noted that, for the $RI = 3$ group, the quadratic model fits the response latency data as well with one independent variable ($MEM \times SEAR$) as the linear model does with two independent variables.

These results are consistent with the expected results with regards to memory set size and search string size. Larger values of each variable resulted in increased response time. However, the regression analysis provided no clarification of whether a linear or a log-linear model best describes the performance-latency data (Briggs, 1974). In the aggregate, the subjects with the longer retention interval had lower error rates. However, at the highest memory loads, the error rates of the two groups were equivalent, but response latencies were longer for the 3-second group than those for the 0-second retention group.

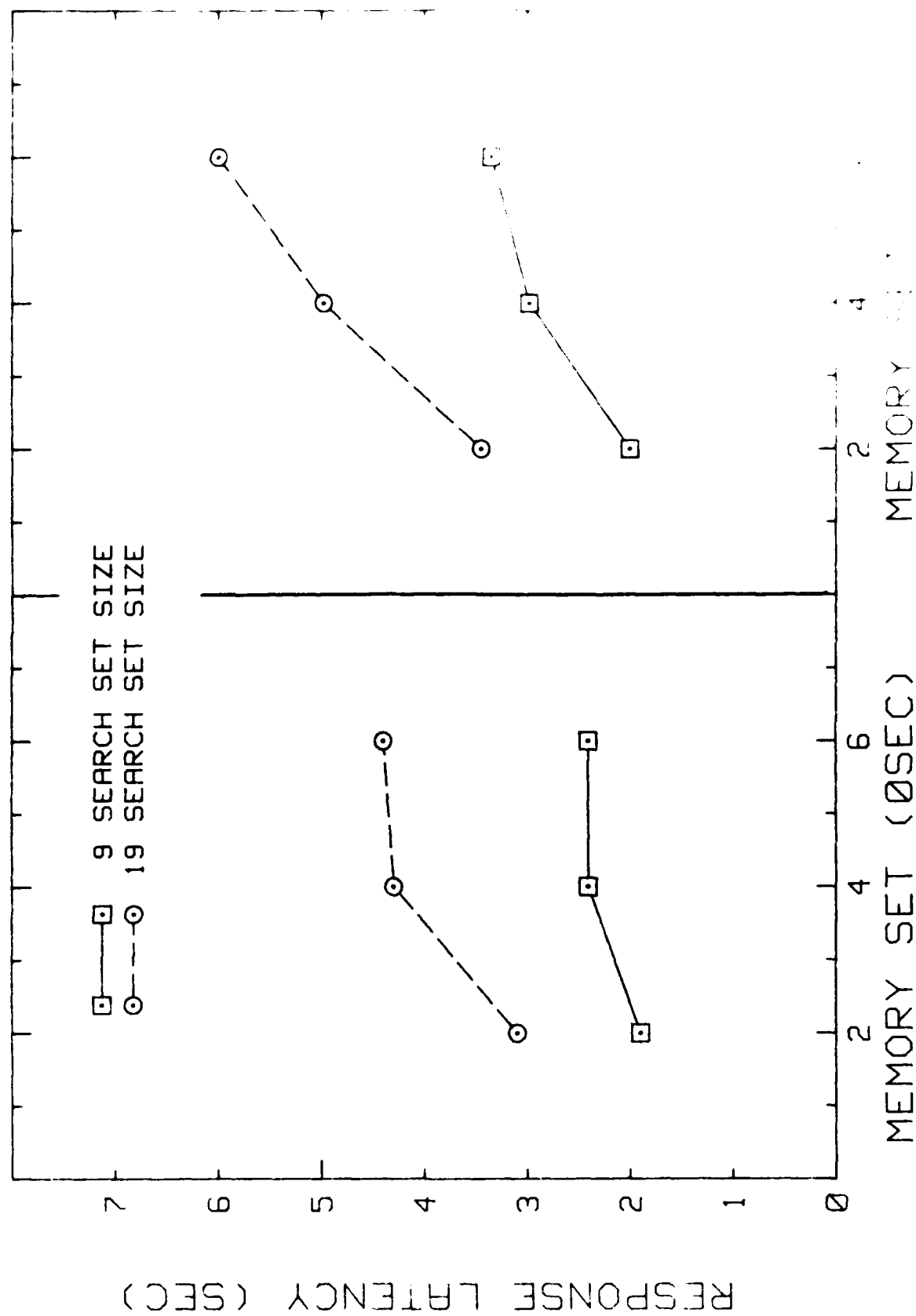


Figure 1. Memory/visual search task: Mean response latency as a function of memory set size and target search set size. The left panel shows data for the first retention interval group; the right panel data for the second.

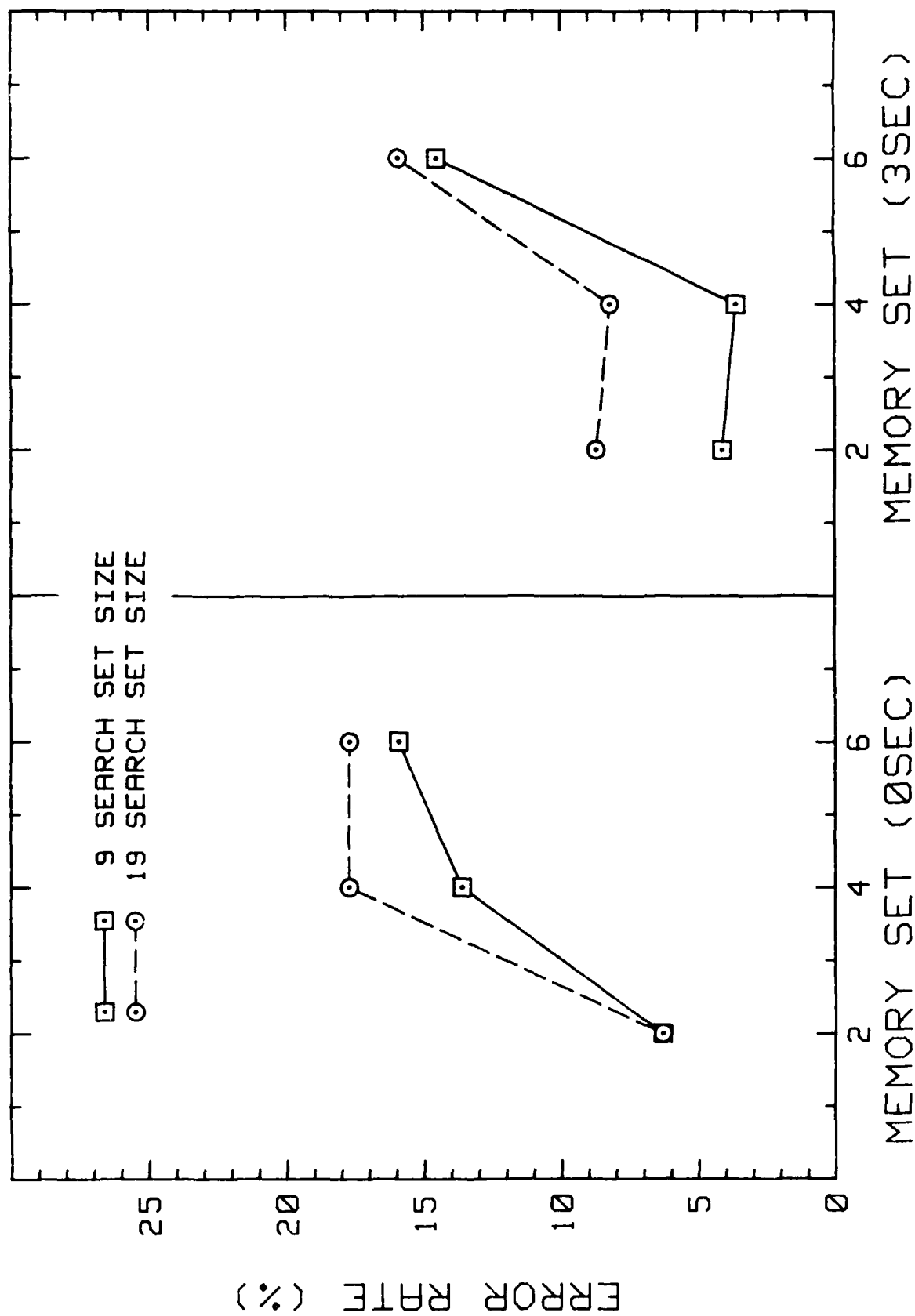


Figure 2. Memory/visual search task: Error rates as a function of memory set size and target search size. The left panel contains data from the 0-second retention interval group; the right panel data from the 3-second group.

Table 1

Regression Coefficients for the Memory Search Task:
Predicting Response Latency (CRT) From Task Variables

MODEL 1: LINEAR (Possible independent variables are memory, search, and block.)

	Retention Interval	
	0 seconds	3 seconds
MEMORY	.24	.51
SEARCH	.15	.20
CONSTANT	.09	-1.11
R^2	.25	.44

MODEL 2: QUADRATIC (Possible independent variables are memory, search, memory*search, memory², search², block.)

	Retention Interval	
	0 seconds	3 seconds
SEARCH ²	.003	
MEMORY*SEARCH	.017	.042
CONSTANT	1.580	1.440
R^2	.25	.44

MODEL 3: LOGARITHMIC (Possible independent variables are log₂ (memory), log₂ (search), block.)

	Retention Interval	
	0 seconds	3 seconds
LOG(MEMORY)	.638	1.282
LOG(SEARCH)	1.410	1.896
CONSTANT	-3.230	-5.640
R^2	.25	.45

Note. Data for the two retention interval groups (RI = 0 and RI = 3 seconds) were analyzed separately.

EXPERIMENT II

Mental arithmetic has been used to assess workload in several previous batteries (see Ogden et al., 1979, for review). Although simple arithmetic is a skill mastered by most literate adults, the method by which we compute is not well-understood. Groen and Parkman (1972) proposed a counting model to explain response latency to simple addition problems. However, Ashcroft and Battaglia (1978) reported research that more closely matched a memory search and retrieval explanation. They found that subjects could determine the veracity of a sum in approximately 900 msec, while the size of the actual digits to be added contributed little (less than 2 msec/digit) to the total reaction times for adult subjects. They did report that response times increased as the square of the digit sum and hence that the total reaction time would increase nonlinearly with the magnitude of the problem.

The mental arithmetic task used in the present study required that the subject calculate a running total of addition and subtraction operations on a series of single digit numbers. This task varied the magnitude of the problem by altering the number of digits in the series. A linear increase in response time would be predicted if including more computations in a single problem merely multiplied the number of operations that had been necessary to compute a two-element problem. However, if computation time was increased as an exponential scale of task difficulty, then the resulting response latencies would increase as the quadratic function described by Ashcroft and Battaglia (1978).

In addition, the stimulus duration of each digit in the problem was varied to determine whether computation on the subtotal could be conducted during the course of the series presentation. Gopher and Sanders (1984) have discussed the question of operation interruption in the context of dual-task time-sharing. They have noted that at some stages, two tasks cannot be time-shared because one stage will essentially capture processing capacity for the duration of its execution. In the analogous aspect of the current task, the question to be tested was whether computation on the partial sum could be accomplished efficiently during the period prior to the presentation of the next digit in the problem. If the subject could perfectly time-share the processing of the current subtotal and the processing of the next incoming stimulus digit, then reaction time (measured from the offset of the final digit) would show a decrease equal to the extra processing time allowed by longer stimulus durations. Stimulus durations were made short enough to interrupt the time required to complete a two-digit computation in order to require interruption of the computation by the presentation of the next digit in the problem.

METHODS

Subjects. Twenty university subjects were tested. Ten of these participants were females and ten were males. Half were graduate students. Each had normal or corrected to normal visual acuity. Each subject participated in an individual session during which he or she completed all conditions of the experiment.

Design. The task levels were the following: (2, 3, or 4 numbers to be totaled) x (250, 400, or 650 msec per number and operant). Each of these nine conditions was presented in two separate blocks of 10 trials. The order of presentation of the nine conditions was counterbalanced as described in Experiment I.

Procedure. The procedure was identical to that used in the first experiment except for the subject's task. In this experiment the subject's task was to perform a serial addition/subtraction task. Randomly selected digits were presented one at a time at the fixation point. In each trial, two digits were displayed in sequence, followed by a plus or a minus sign. If the trial was in a three-item condition, a third digit and another plus or minus sign followed. If it was a four-digit trial, a fourth digit and sign followed. The subject's task was to perform this series of computations and to enter the units digit of the total (e.g., 8, 6, + yields 14, so enter 4). If a negative answer resulted, the subject added 10 and entered the resulting units digit. The duration of each digit and each operant sign was constant within a block of 10 trials. Between trials, this duration varied as 250, 400, and 650 msec.

The modified test battery provided the opportunity to vary (via menu) the following parameters in this serial arithmetic task: the number of trials in each block, the number of digits, and the duration of the digits and the arithmetic operants.

RESULTS

The mean response times to correct trials in each block were analyzed in an analysis of variance. A mixed-factors design was used to determine the significance of the three repeated measures (number of items, duration of items, and trial blocks). The mean response times as a function of number of items and item duration are graphed in Figure 3. The effect of number of items was significant, $F(2,38) = 36.38$, $p < .001$. Practice (blocks), $F(1,19) = 26.35$, $p < .001$, as well as the interaction of Item x Block, $F(2,38) = 4.19$, $p < .05$, were found to be significant. Response speed improved by 450 msec between blocks as a function of practice; however, the conditions with more items improved more. When the response latency is considered as cost per digit, the information processing rate is virtually constant in all Block 1 conditions (595, 583, 555 msec for two, three, and four digits, respectively). Furthermore, this value is decreased by a constant in the second Block to 449, 437, and 397 msec for two, three, and four digits, respectively. Item duration did not affect response latency (see Figure 3).

Error rates increased as the number of stimulus items increased. This effect appears to be stronger for the shorter duration stimuli. The error rates as a function of number and duration of items were graphed in Figure 4. In addition, the effect of practice improved the error rate from 12.6 percent to 9 percent between Blocks 1 and 2. This practice effect did not appear to interact with either number of items or duration.

While both latency and accuracy showed a practice effect and reflected variation in number of items, only error rate was affected by changes in item

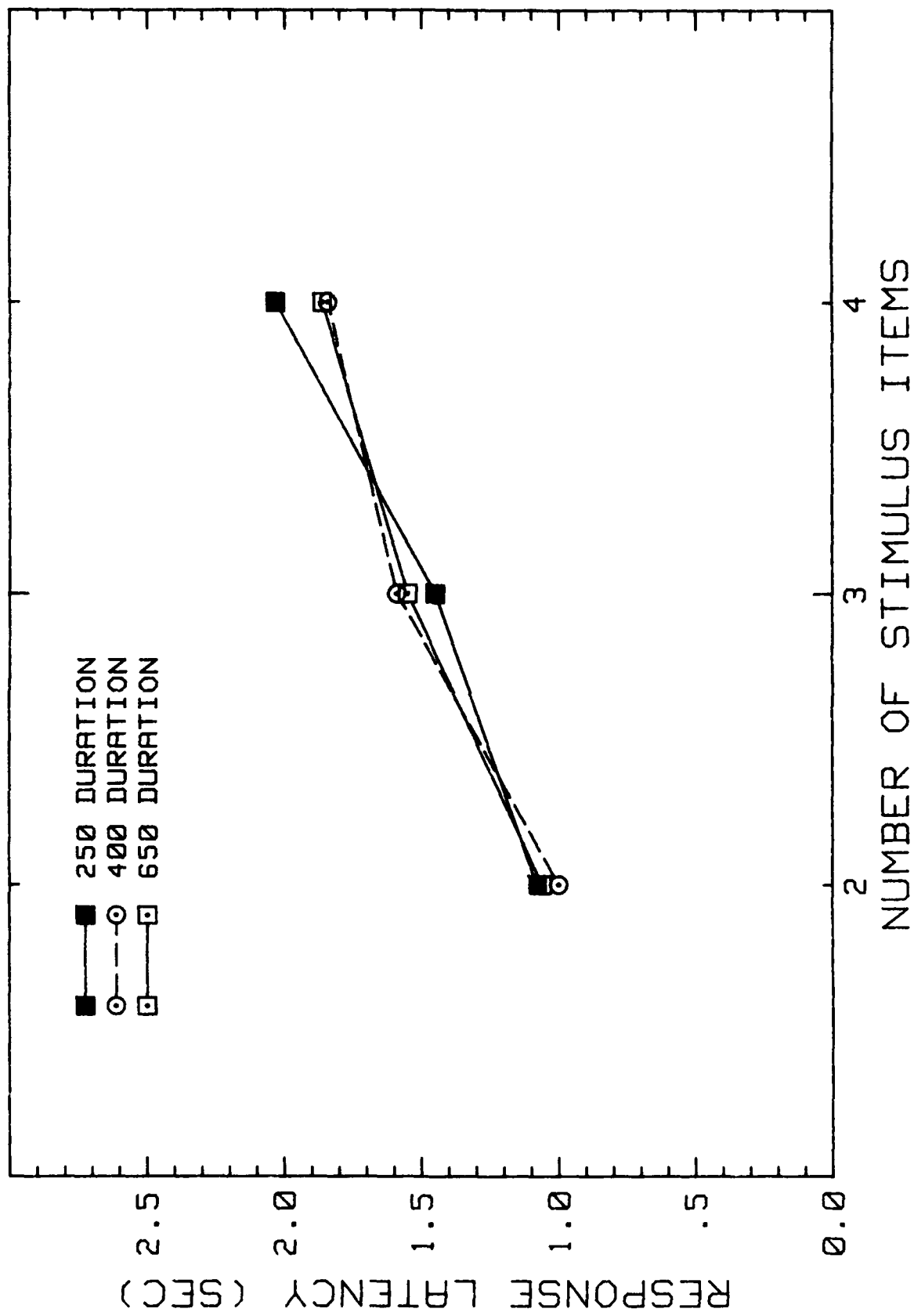


Figure 3. Add/subtract task: Mean response latency as a function of number of stimulus items and the stimulus durations.

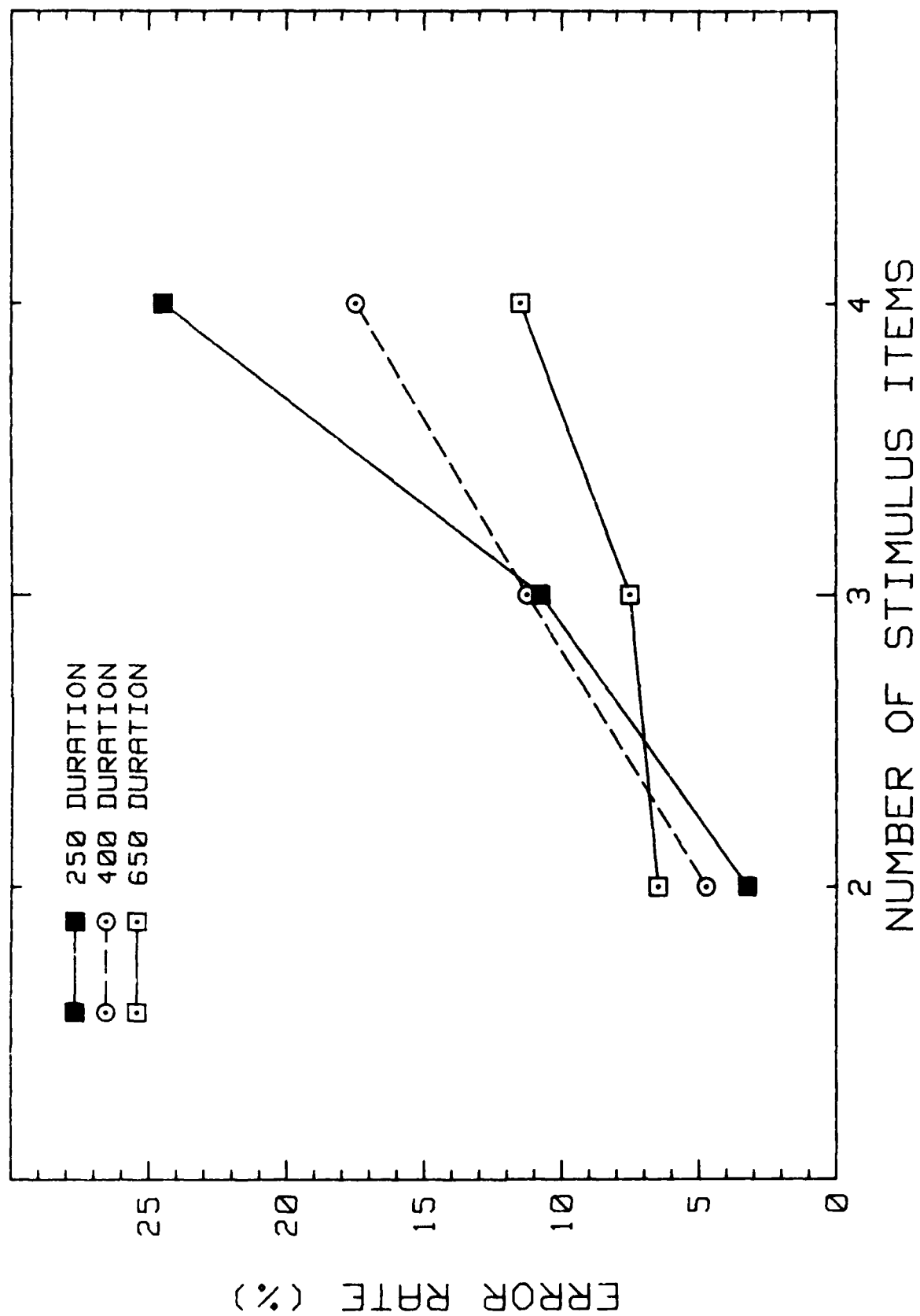


Figure 4. Add/subtract task: Error rates as a function of number of stimulus items and the stimulus durations.

duration. Although the total amount of time from the presentation of the first stimulus to the response cue varied from 700 msec to 1,700 msec for trials with only two digits and from 1,750 to 4,500 msec for trials with four digits, response latency was unaltered by this duration variable.

The data for latencies of correct responses were fit to linear, quadratic, and logarithmic models. The best linear model entered both item and block. Both the quadratic and the logarithmic models also entered these two variables; however, the logarithmic fit was less satisfactory. The regression equations for each model are listed in Table 2.

Varying duration affected error rates, but it did not affect response latency. This separation of the duration effect on the two independent variables is consistent with the following interpretation: Subjects waited until the entire list of digits was complete before they began to compute the sum. Because the shorter presentation rate (250 msec) allowed them less time to consolidate each item (resolving the physical image and perhaps attaching a verbal name), error rates were higher for shorter stimulus presentation durations.

EXPERIMENT III

In the third experiment subjects were asked to respond to an arithmetic task in which the stimuli were present for the entire computation. In this way, a memory component for the stimulus numbers was not necessary, but only a computational memory component (for carrying values from column to column). The task conditions varied in difficulty as a function of the size (units, tens, hundreds) and the number of items to be added. Ashcraft and Stazyk (1981) reported that increasing problem size from one-digit to two-digit addends increased response time sharply from 959 to 1,378 msec. They suggested that response latency should reflect the difficulty in the summing of each column plus a constant amount of time for each carry operation. In the present experiment, both the number of digits per addend and the number of addends in the problem were varied to determine whether the increase in problem difficulty from these two sources produced equal increments in response latency.

METHODS

Subjects. Twenty university subjects were tested. Ten of these participants were females and ten were males. Half were graduate students. All were found to have normal or corrected to normal visual acuity. Each subject was tested in an individual session during which he or she completed all conditions of the experiment.

Design. The task levels were the following: 2 and 3 addends (1 and 2 x (1, 2, or 3 digits in each number). Each of these six arithmetic tasks was presented in two separate blocks of 10 trials. The order of presentation of the conditions was counterbalanced as described in Experiment 1.

Table 2

Regression Coefficients for the Serial 7 Arithmetic Task:
Predicting Response Latency (CRI) From Task Variables

MODEL 1: LINEAR (Possible independent variables are duration, item, block.)

ITEM	.432
BLOCK	-.454
CONSTANT	.879
R^2	.21

MODEL 2: QUADRATIC (Possible independent variables are duration, item, duration*item, item², duration², block.)

SAME AS LINEAR MODEL

MODEL 3: LOGARITHMIC (Possible independent variables are log₂(duration), log₂(item), block.)

LOG(ITEM)	.860
BLOCK	.454
CONSTANT	.858
R^2	.21

Procedure. The procedure was identical to that used in the first experiment except for the subject's task. In this experiment the subject's task was to perform an addition task. Numbers were presented simultaneously in column format in the center of the screen. In each block of trials, either two or three numbers were to be added. In each block the digits in each number were fixed at one, two, or three. In each trial, the numbers were displayed until the subject entered the first digit of the total; then the column of numbers disappeared.

The modified test battery provided the opportunity to vary (via menu) the following parameters in this addition task: the number of trials in each block, the number of digits in each item, and the number of items to be added.

RESULTS

The mean response times to correct trials in each block were analyzed in an analysis of variance. A repeated-measures design was used to test the effect of number of items to be added, number of digits in each item, and trial block. The effect of blocks was not found to be significant in any main effect or interaction; therefore, the data were graphed for the average of the two blocks of trials in each condition.

The mean response times for conditions were graphed in Figure 5. The effects of items, $F(1,19) = 119.4$, digits, $F(2,38) = 187.4$, and their interaction (Item x Digit), $F(2,38) = 61.8$, were all significant at $p < .001$.

Error rates increased as the number of digits in each addend increased (1 vs. 2 vs. 3). While the error rates for adding two or three one-digit numbers were virtually identical, error rates for larger addends (two- and three-digit numbers) were higher for three-item problems than for two-item problems. The error rates were graphed in Figure 6.

When the graphs of the two dependent variables are considered together, both show decrements in performance as the number of items and digits increased.

The data for latencies of correct responses were fit to linear, quadratic, and logarithmic models for each retention interval group. The best linear model entered both number of items and number of digits. The quadratic and the logarithmic model provide a similar fit by adding different functions of the same two independent variables. The regression equations for each model are listed in Table 3.

All three models do a very good job of fitting the data. The best model is the quadratic model, which fits the data with one independent variable (ITEM * DIGIT) and explains the greatest portion of the variance ($R^2 = .74$).

Both error rates and response latencies were affected in a very systematic way by the increasing load of computations. Increasing the number of items to be added or the number of digits within each item increased latency by approximately the same amount (4.70 and 4.72 seconds for each unit increase). The cost to the error rate of increasing the size of the computation was about 8 percent per unit.

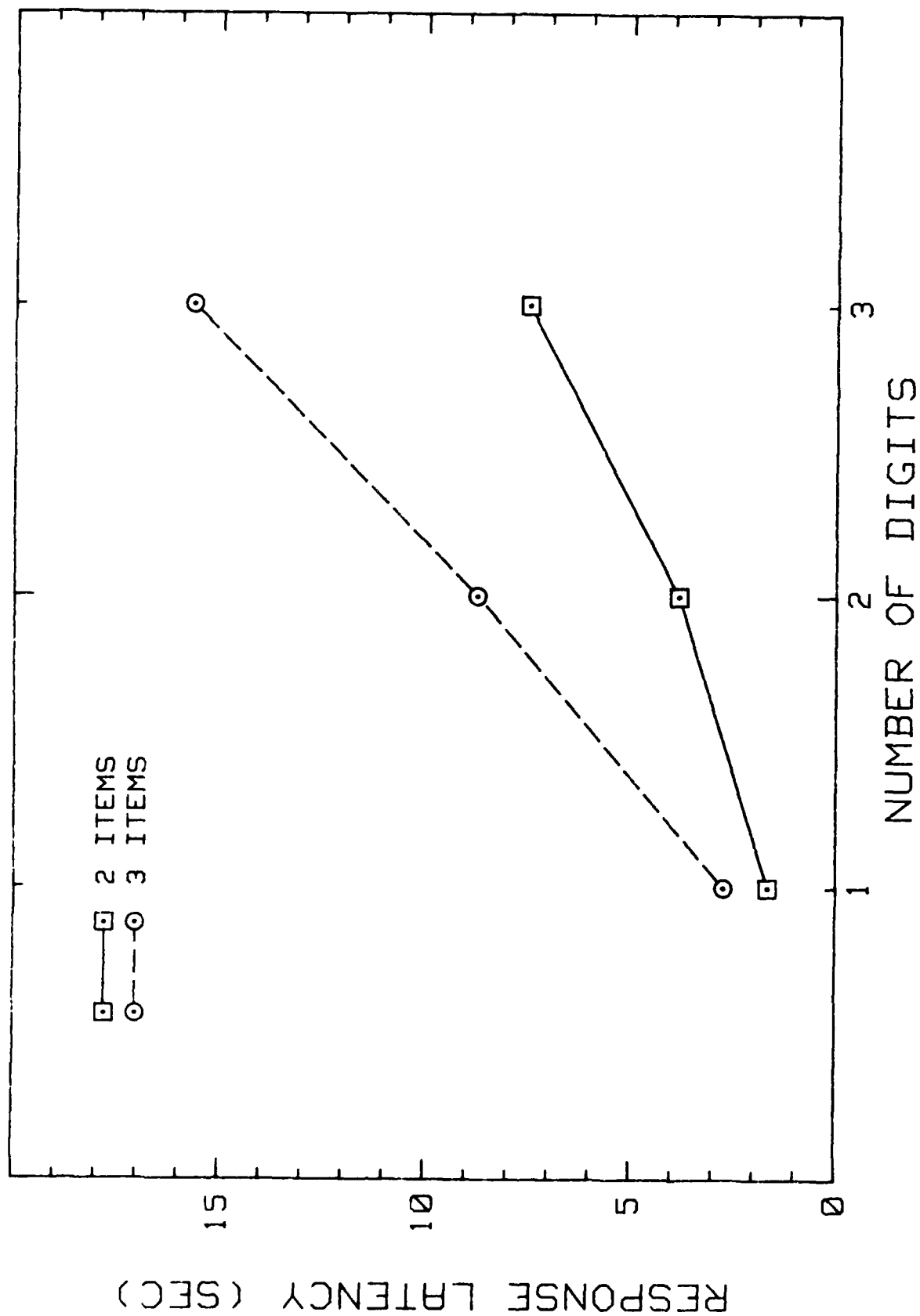


Figure 5. Column addition task: Mean response latency as a function of number of digits per item and number of items to be added.

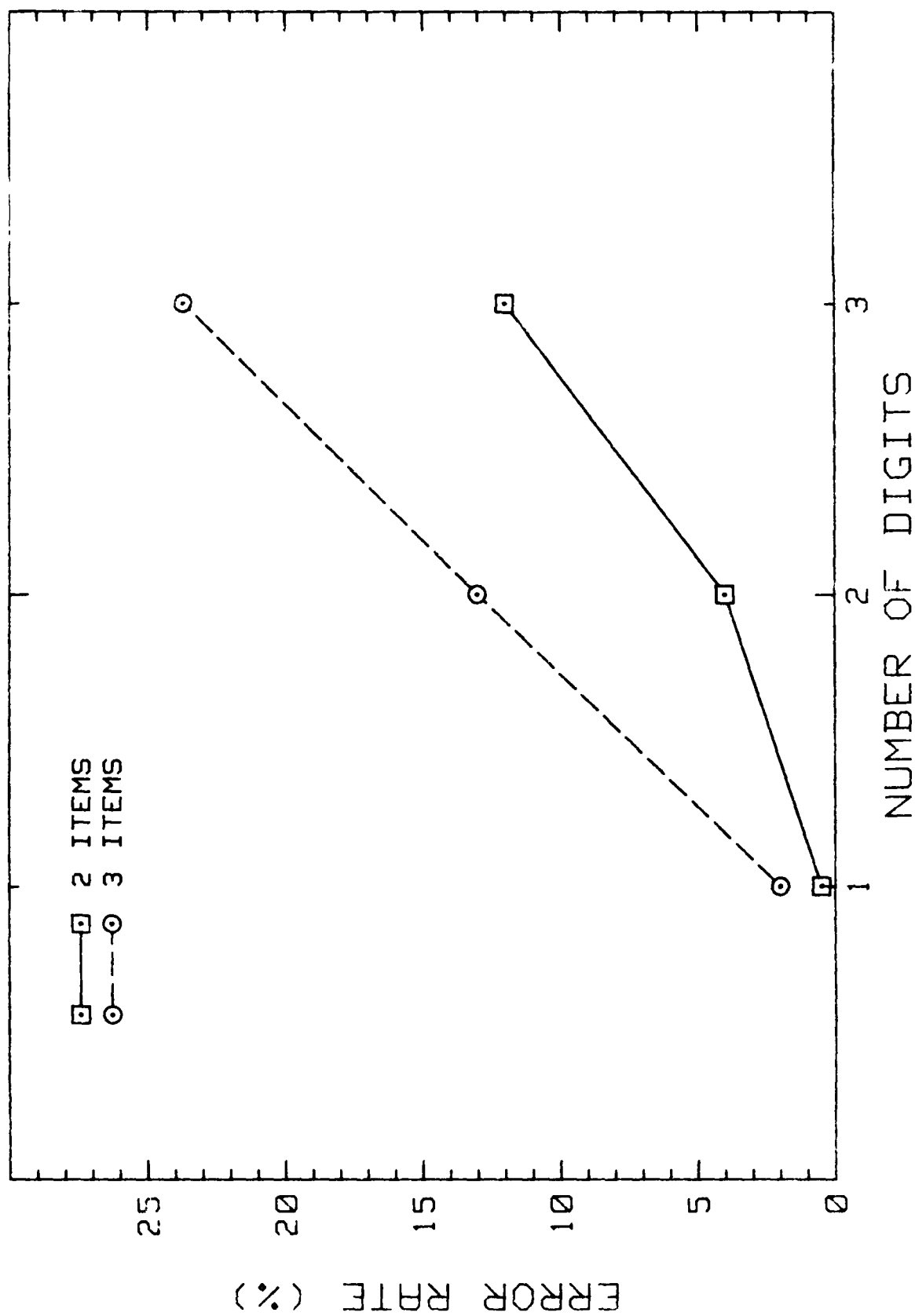


Figure 6. Column addition task: Error rates as a function of number of digits per item and number of items to be added.

Table 3

Regression Coefficients for the Addition Task:
Predicting Response Latency (CRT) From Task Variables

MODEL 1: LINEAR (Possible independent variables are digit, item, block.)

DIGIT	4.70
ITEM	4.72
CONSTANT	-14.52
R^2	.69

MODEL 2: QUADRATIC (Possible independent variables are digit, item, digit*item, digit², item², block.)

ITEM*DIGIT	3.42
R^2	.74

MODEL 3: LOGARITHMIC (Possible independent variables are $\log_2(\text{digit})$, $\log_2(\text{item})$, block.)

LOG(DIGIT)	5.74
LOG(ITEM)	8.07
CONSTANT	-8.69
R^2	.66

GENERAL DISCUSSION

In a series of three experiments, the subjects' tasks were varied to alter the difficulty of the task. Performance measures of accuracy and speed were measured to determine the influence of experimenter-defined difficulty (i.e., task parameter) on the subject's performance. All three tasks required a memory load and two of the tasks required a computational load. The purpose of the research was to establish the sensitivity of performance measures to variations in task difficulty. The pattern of performance measures was examined to determine the effects of the various levels of memory and computational load.

In the range of values tested, all task parameters, except stimulus duration in the serial arithmetic (addition/subtraction) task (Experiment II), affected performance, showing a systematic increase in latency and error rates as a function of task difficulty. In this one exception, latency was not sensitive to stimulus duration even over a large range of values (750 to 1,950 msec for the smallest number of computations/problem; and 1,750 to 4,550 msec for the largest number of computations/problem). It appears either (1) subjects were doing no computations during the serial presentation of the items to be added, or (2) subjects had to recalculate some portion of the computations after the final operation (beginning of the response interval). In either case, the net result was that the response latency was unaffected by the amount of time that the stimulus items had been presented. However, error rates were greater when the subject had had less time to view each item and operant. The improvement in accuracy as a function of viewing time may be a function of greater time to analyze and label the stimulus before having to process the next incoming item.

The results of the memory/visual search task (Experiment I) were complex because of the possible speed/accuracy trade-off between the two retention interval groups. The most parsimonious explanation of the differences in performance between the two groups is that the 0-second retention interval group adopted a higher accuracy criterion at the cost of response speed. However, further examination of the patterns of responding lead to a more interesting (though post hoc) hypothesis. Error rates were comparable for the two groups at the longest (6) memory set size; however, response latencies were slower for the group with the longer retention interval. Consolidation of the memory set items into a different (perhaps verbal) code during the extended retention interval would necessitate a different retrieval strategy and a longer search to match time (Klatzky, 1983).

Latency in the column addition task (Experiment III) increased as a function of the total content in the stimulus; i.e., adding three 2-digit items was as time-consuming as adding two 3-digit numbers. This can be seen in either the linear or the quadratic regression model. Either model explains a sufficient portion of the variance ($R^2 = .69$ and $.74$, respectively) to be considered a satisfactory predictor (Cunniff, 1975). However, differentiating between the linear and the quadratic predictor models is not possible from these data. Future research would be needed to determine whether it is the total number of arithmetic operations (the quadratic model) or the independent influences of items and digits (the linear model) that influence response latency. The coefficients in the linear model are $.11$ and $.122$ seconds for digits and items, respectively; in the quadratic model the $2nd$ variable term is the product of item*digit. Thus the two variables can be used interchangeably to predict response latency.

In conclusion, performance was found to be a sensitive indicator of changes in the task parameters for the three tasks tested. The test battery is an easy, economical method of administering variations of these tasks. Future research measuring performance, evoked potentials, and subjective responses is planned using this battery of three tasks.

REFERENCES

- Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. Journal of Experimental Psychology: Human Learning and Memory, 4, 527-538.
- Ashcraft, M. H., & Staszko, E. H. (1981). Mental addition: A test of three verification models. Memory and Cognition, 9, 185-196.
- Briggs, G. E. (1974). On the predictor variable for choice reaction time. Memory and Cognition, 2, 575-580.
- Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21, 515-528.
- Folkard, S., Knauth, P., Monk, T. H., & Rutenfranz, J. (1976). The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system. Ergonomics, 19, 479-488.
- Gopher, D., & Sanders, A. F. (1984). S-Oh-R: Oh Stages! Oh Resources! In Cognition and motor processes. (Ed. Prinz, W., & Sanders, A.F.), Springer-Verlag: Berlin, 231-253.
- Groen, C. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. Psychological Review, 79, 329-343.
- Kaplan, I. T., Carvellas, T., & Metlay, W. (1966). Visual search and immediate memory. Journal of Experimental Psychology, 71, 488-493.
- Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1, 153-161.
- Klatzky, R. L. (1984). Memory and awareness: An information-processing perspective. New York: W. H. Freeman and Company.
- Knowles, W. B. (1963). Operator loading tasks. Human Factors, 5, 155-161.
- LePlat, J. (1978). Factors determining work-load. Ergonomics, 21, 143-149.
- Moray, N. (1982). Subjective mental workload. Human Factors, 24, 25-40.
- Neisser, U., Novick, R., & Lazar, R. (1963). Searching for ten targets simultaneously. Perceptual and Motor Skills, 17, 955-961.
- Nunnally, J. C. (1975). Psychometric Theory. (2nd edition). New York: McGraw-Hill.
- Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58, 193-198.
- SPSS: Statistical Package for the Social Sciences. (1970, 2nd ed.). New York: McGraw-Hill.
- Simmons, R. R. (1979). Methodological considerations of visual workloads of helicopter pilots. Human Factors, 21, 353-367.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.
- Thorne, D. R., Genser, S. G., Sing, H. C., & Hegge, F. W. (1985). The Walter Reed Performance Assessment Battery. Neurobehavioral Toxicology and Teratology, 7, 415-418.
- Weiner, J. S. (1982). The measurement of human workload. Ergonomics, 25, 953-965.
- Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, Ohio: C. E. Merrill.
- Wickens, C. D., & Kramer, A. (1985). Engineering psychology. Annual Review of Psychology, 36, 307-348.

APPENDIX
ANALYSIS OF VARIANCE TABLES

EXPERIMENT 1: ANALYSIS OF VARIANCE TABLE

	SOURCE	SS	MS	df	F
BETWEEN S	RETENTION INTER	24.41	24.41	1	1.18
		373.24	20.74	18	
WITHIN S	BLOCKS	.03	.03	1	.04
	RET INT X BLOCK	.09	.09	1	.13
		11.87	.66	18	
	MEMORY SET SIZE	194.36	194.36	1	201.85
	RET INT X MEMORY	4.70	4.70	1	4.9
		17.33	.96	18	
	SEARCH SIZE	87.32	43.66	2	25.43
	RET INT X SEARCH	9.19	4.59	2	2.68
		61.82	1.72	36	
	BLOCK X MEMORY	.49	.49	1	1.39
	RET X BL X MEM	.01	.01	1	.01
		6.37	.35	18	
	BLOCK X SEARCH	1.70	.85	2	2.54
	RET X BL X SEAR	.51	.26	2	.76
		12.08	.34	36	
	MEM X SEARCH	6.40	3.20	2	20.83
	RET X MEM X SEAR	1.13	.56	2	3.7
		5.52	.15	36	
	BL X MEM X SEAR	.26	.13	2	.66
	RET X B X M X S	.11	.06	2	.28
		7.12	.20	36	
	TOTAL	826.07		239	

EXPERIMENT II: ANALYSIS OF VARIANCE TABLE

SOURCE	SS	MS	df	F
BLOCKS	18.53	18.50	1	26.43
	13.36	.70	19	
DURATION	.13	.06	2	.26
	8.92	.23	38	
ITEMS	45.02	22.51	2	36.31
	23.51	.62	38	
BLOCK X DUR	1.30	.65	2	1.51
	16.47	.43	38	
BLOCK X ITEMS	1.69	.84	2	4.20
	7.67	.20	38	
DUR X ITEMS	1.32	.33	4	2.36
	10.52	.14	76	
BL X DUR X ITEMS	.76	.19	4	.79
	18.29	.24	76	
BET S ERR	132.76	6.99	19	
TOTAL	300.25		359	

EXPERIMENT III: ANALYSIS OF VARIANCE TABLE

SOURCE	SS	MS	df	F
BLOCKS	3.39 39.46	3.39 2.08	1 19	1.63
ITEMS	1338.59 212.98	1338.59 11.21	1 19	119.42
DIGITS	3556.84 360.58	1778.42 9.49	2 38	187.42
BLOCK X ITEMS	1.76 46.81	1.76 2.46	1 19	.71
BLOCK X DIGITS	1.95 77.32	.98 2.03	2 38	.48
ITEM X DIGIT	504.47 155.00	252.23 4.08	2 38	61.84
BL X ITEM X DIG	4.35 82.50	2.18 2.17	2 38	1
BETWEEN S ERR	721.79	37.99	19	
TOTAL	7107.79		239	

END

DATE

FILMED

JAN

1988